

SEQUENCE LISTING

Sub 31
~~(1) GENERAL INFORMATION:~~

~~(i) APPLICANT: De Robertis, Edward M.
Bouwmeester, Tewis~~

~~(ii) TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
Factors~~

~~(iii) NUMBER OF SEQUENCES: 10~~

~~(iv) CORRESPONDENCE ADDRESS:~~

~~(A) ADDRESSEE: Majestic, Parsons, Siebert & Hsue
(B) STREET: Four Embarcadero Center, Suite 1100
(C) CITY: San Francisco
(D) STATE: California
(E) COUNTRY: U.S.A.
(F) ZIP: 94111-4106~~

~~(v) COMPUTER READABLE FORM:~~

~~(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patentin Release #1.0, Version #1.25~~

~~(vi) CURRENT APPLICATION DATA:~~

~~(A) APPLICATION NUMBER: US 08/878,474
(B) FILING DATE: 18-JUN-1997
(C) CLASSIFICATION:~~

~~(vii) PRIOR APPLICATION DATA:~~

~~(A) APPLICATION NUMBER: US 60/020,150
(B) FILING DATE: 20-JUN-1996~~

~~(viii) ATTORNEY/AGENT INFORMATION:~~

~~(A) NAME: Siebert, J. Suzanne
(B) REGISTRATION NUMBER: 28,758
(C) REFERENCE/DOCKET NUMBER: 3100.002US1~~

~~(ix) TELECOMMUNICATION INFORMATION:~~

~~(A) TELEPHONE: 415/248-5500
(B) TELEFAX: 415/362-5418~~

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 270 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Leu Leu Asn Val Leu Arg Ile Cys Ile Ile Val Cys Leu Val Asn
1 5 10 15

Asp Gly Ala Gly Lys His Ser Glu Gly Arg Glu Arg Thr Lys Thr Tyr
20 25 30

Ser Leu Asn Ser Arg Gly Tyr Phe Arg Lys Glu Arg Gly Ala Arg Arg
35 40 45

Ser Lys Ile Leu Leu Val Asn Thr Lys Gly Leu Asp Glu Pro His Ile
50 55 60

Gly His Gly Asp Phe Gly Leu Val Ala Glu Leu Phe Asp Ser Thr Arg
65 70 75 80

Thr His Thr Asn Arg Lys Glu Pro Asp Met Asn Lys Val Lys Leu Phe
85 90 95

Ser Thr Val Ala His Gly Asn Lys Ser Ala Arg Arg Lys Ala Tyr Asn
100 105 110

Gly Ser Arg Arg Asn Ile Phe Ser Arg Arg Ser Phe Asp Lys Arg Asn
115 120 125

Thr Glu Val Thr Glu Lys Pro Gly Ala Lys Met Phe Trp Asn Asn Phe
130 135 140

Leu Val Lys Met Asn Gly Ala Pro Gln Asn Thr Ser His Gly Ser Lys
145 150 155 160

Ala Gln Glu Ile Met Lys Glu Ala Cys Lys Thr Leu Pro Phe Thr Gln
165 170 175

Asn Ile Val His Glu Asn Cys Asp Arg Met Val Ile Gln Asn Asn Leu
180 185 190

Cys Phe Gly Lys Cys Ile Ser Leu His Val Pro Asn Gln Gln Asp Arg
195 200 205

Arg Asn Thr Cys Ser His Cys Leu Pro Ser Lys Phe Thr Leu Asn His
210 215 220

Leu Thr Leu Asn Cys Thr Gly Ser Lys Asn Val Val Lys Val Val Met
 225 230 235 240
 Met Val Glu Glu Cys Thr Cys Glu Ala His Lys Ser Asn Phe His Gln
 245 250 255
 Thr Ala Gln Phe Asn Met Asp Thr Ser Thr Thr Leu His His
 260 265 270

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1411 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAATTCTAA AAGCGGCACA GTGCAGGAAC AGCAAGTCGC TCAGAACAC TGCAGGGTCT	60
AGATATCATA CAATGTTACT AAATGTACTC AGGATCTGTA TTATCGTCTG CCTTGTGAAT	120
GATGGAGCAG GAAAACACTC AGAAGGACGA GAAAGGACAA AAACATATTC ACTAACAGC	180
AGAGGTTACT TCAGAAAAGA AAGAGGAGCA CGTAGGAGCA AGATTCTGCT GGTGAATACT	240
AAAGGTCTTG ATGAACCCC AATTGGCAT GGTGATTTG GCTTAGTAGC TGAACATATT	300
GATTCCACCA GAACACATAC AAACAGAAAA GAGCCAGACA TGAACAAAGT CAAGCTTTG	360
TCAACAGTTG CCCATGGAAA CAAAAGTGCAGA AGAAGAAAAG CTTACAATGG TTCTAGAAGG	420
AATATTTTTT CTCGCCGTTC TTTTGATAAA AGAAATACAG AGGTTACTGA AAAGCCTGGT	480
GCCAAGATGT TCTGGAACAA TTTTTGGTT AAAATGAATG GAGCCCCACA GAATACAAGC	540
CATGGCAGTA AAGCACAGGA AATAATGAAA GAAGCTTGCA AAACCTTGCC CTTCACTCAG	600
AATATTGTAC ATGAAAATG TGACAGGATG GTGATACAGA ACAATCTGTG CTTGGTAAA	660
TGCATCTCTC TCCATGTTCC AAATCAGCAA GATCGACGAA ATACTGTTTC CCATTGCTTG	720
CCGTCCAAAT TTACCCCTGAA CCACCTGACG CTGAATTGTA CTGGATCTAA GAATGTAGTA	780
AAGGTTGTCA TGATGGTAGA GGAATGCACG TGTGAAGCTC ATAAGAGCAA CTTCCACCAA	840

ACTGCACAGT TTAACATGGA TACATCTACT ACCCTGCACC ATTAAAAGGA CTGTCTGCCA 900
 TACAGTATGG AAATGCCAT TTGTTGGAAT ATTGTTACA TGCTATGTAT CTAAGCATT 960
 ATGTTGCCTT CTGTTTCATA TAACCACATG GAATAAGGAT TGTATGAATT ATAATTAACA 1020
 AATGGCATT TGTGTAACAT GCAAGATCTC TGTTCCATCA GTTGCAAGAT AAAAGGCAAT 1080
 ATTTGTTGA CTTTTTCTA CAAATGAAT ACCCAAATAT ATGATAAGAT AATGGGGTCA 1140
 AAACTGTTAA GGGGTAATGT AATAATAGGG ACTAACAAACC AATCAGCAGG TATGATTAC 1200
 TGGTCACCTG TTTAAAAGCA AACATCTTAT TGTTGCTAT GGGTTACTGC TTCTGGCAA 1260
 AATGTGTGCC TCATAGGGGG GTTAGTGTGT TGTGTACTGA ATTAATTGTA TTTATTCAT 1320
 TGTTACAATG AAGAGGATGT CTATGTTAT TTCACTTTA TTAATGTACA ATAAATGTTCA 1380
 TTGTTCTTT AAAAAAAA AAAAATCGA G 1411

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 318 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Ser Arg Thr Arg Lys Val Asp Ser Leu Leu Leu Leu Ala Ile Pro 15
 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
 Gly Leu Ala Leu Leu Leu Pro Asn Ala Tyr Cys Ala Ser Cys Glu
 20 25 30
 Pro Val Arg Ile Pro Met Cys Lys Ser Met Pro Trp Asn Met Thr Lys
 35 40 45
 Met Pro Asn His Leu His His Ser Thr Gln Ala Asn Ala Ile Leu Ala
 50 55 60
 Ile Glu Gln Phe Glu Gly Leu Leu Thr Glu Cys Ser Gln Asp Leu
 65 70 75 80
 Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys Thr Ile Asp Phe

Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys Glu Arg Ala Arg
 100 105 110
 Ala Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His Thr Trp Pro Glu
 115 120 125
 Ser Leu Ala Cys Glu Glu Leu Pro Val Tyr Asp Arg Gly Val Cys Ile
 130 135 140
 Ser Pro Glu Ala Ile Val Thr Val Glu Gln Gly Thr Asp Ser Met Pro
 145 150 155 160
 Asp Phe Ser Met Asp Ser Asn Asn Gly Asn Cys Gly Ser Gly Arg Glu
 165 170 175
 His Cys Lys Cys Lys Pro Met Lys Ala Thr Gln Lys Thr Tyr Leu Lys
 180 185 190
 Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Val Lys Val
 195 200 205
 Lys Cys His Asp Ala Thr Ala Ile Val Glu Val Lys Glu Ile Leu Lys
 210 215 220
 Ser Ser Leu Val Asn Ile Pro Lys Asp Thr Val Thr Leu Tyr Thr Asn
 225 230 235 240
 Ser Gly Cys Leu Cys Pro Gln Leu Val Ala Asn Glu Glu Tyr Ile Ile
 245 250 255
 Met Gly Tyr Glu Asp Lys Glu Arg Thr Arg Leu Leu Leu Val Glu Gly
 260 265 270
 Ser Leu Ala Glu Lys Trp Arg Asp Arg Leu Ala Lys Lys Val Lys Arg
 275 280 285
 Trp Asp Gln Lys Leu Arg Arg Pro Arg Lys Ser Lys Asp Pro Val Ala
 290 295 300
 Pro Ile Pro Asn Lys Asn Ser Asn Ser Arg Gln Ala Arg Ser
 305 310 315

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1875 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCCCTT TCACACAGGA CTCCTGGCAG AGGTGAATGG TTAGCCCTAT GGATTTGGTT 60
TGTTGATTT GACACATGAT TGATTGCTT CAGATAGGAT TGAAGGACTT GGATTTTAT 120
CTAATTCTGC ACTTTAAAT TATCTGAGTA ATTGTTCATT TTGTATTGGA TGGGACTAAA 180
GATAAACTTA ACTCCTTGCT TTTGACTTGC CCATAAACTA TAAGGTGGGG TGAGTTGTAG 240
TTGCTTTAC ATGTGCCAG ATTTCCCTG TATTCCCTGT ATTCCCTCTA AAGTAAGCCT 300
ACACATACAG GTTGGCAGA ATAACAATGT CTCGAACAAG GAAAGTGGAC TCATTACTGC 360
TACTGGCCAT ACCTGGACTG GCGCTTCTCT TATTACCCAA TGCTTACTGT GCTTCGTGTG 420
AGCCTGTGCG GATCCCCATG TGCAAATCTA TGCCATGGAA CATGACCAAG ATGCCAACCC 480
ATCTCCACCA CAGCACTCAA GCCAATGCCA TCCTGGCAAT TGAACAGTTT GAAGGTTGCG 540
TGACCACTGA ATGTAGCCAG GACCTTTGT TCTTCTGTG TGCCATGTAT GCCCCCATT 600
GTACCATCGA TTTCCAGCAT GAACCAATTA AGCCTTGCAA GTCCGTGTGC GAAAGGGCCA 660
GGGCCGGCTG TGAGCCCATT CTCATAAAGT ACCGGCACAC TTGGCCAGAG AGCCTGGCAT 720
GTGAAGAGCT GCCCGTATAT GACAGAGGAG TCTGCATCTC CCCAGAGGCT ATCGTCACAG 780
TGGAAACAAGG AACAGATTCA ATGCCAGACT TCTCCATGGA TTCAAACAAT GGAAATTGCG 840
GAAGCGGCAG GGAGCACTGT AAATGCAAGC CCATGAAGGC AACCCAAAAG ACGTATCTCA 900
AGAATAATTAA CAATTATGTA ATCAGAGCAA AAGTGAAGA GGTGAAAGTG AAATGCCACG 960
ACGCAACAGC AATTGTGGAA GTAAAGGAGA TTCTCAAGTC TTCCCTAGTG AACATTCTA 1020
AAGACACAGT GACACTGTAC ACCAACTCAG GCTGCTTGTG CCCCCAGCTT GTTGCCAATG 1080
AGGAATACAT ATTATGGGC TATGAAGACA AAGAGCGTAC CAGGCTTCTA CTAGTGGAAAG 1140
GATCCTTGGC CGAAAAATGG AGAGATCGTC TTGCTAAGAA AGTCAAGCGC TGGGATCAA 1200
AGCTTCGACG TCCCAGGAAA AGCAAAGACC CCGTGGCTCC AATTCCCAAC AAAAACAGCA 1260
ATTCCAGACA AGCGCGTAGT TAGACTAACG GAAAGGTGTA TGGAAACTCT ATGGACTTTG 1320
AAACTAAGAT TTGCATTGTT GGAAGAGCAA AAAAGAAATT GCACTACAGC ACGTTATATT 1380
CTATTGTTA CTACAAGAAG CTGGTTTAGT TGATTGTAGT TCTCCTTCC TTCTTTTTT 1440

TTATAACTAT ATTTGCACGT GTTCCCAGGC AATTGTTTA TTCAACTTCC AGTGACAGAG 1500
 CAGTGACTGA ATGTCTCAGC CAAAGAAC TCAATTCAATT TCTGATCAAC TAATGGTGAC 1560
 AAGTGTGTTGA TACTTGGGGA AAGTGAACTA ATTGCAATGG TAAATCAGAG AAAAGTTGAC 1620
 CAATGTTGCT TTTCTGTAG ATGAACAAAGT GAGAGATCAC ATTTAAATGA TGATCACCTT 1680
 CCATTTAATA CTTTCAGCAG TTTAGTTAG ATGACATGTA GGATGCACCT AAATCTAAAT 1740
 ATTTTATCAT AAATGAAGAG CTGGTTAGA CTGTATGGTC ACTGTTGGGA AGGTAAATGC 1800
 CTACTTGTC AATTCTGTT TAAAAATTGC CAAATAAAAT ATTAAGTCCT AAATAAAAAA 1860
 AAAAAAAA AAAAAA 1875

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 979 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Leu Leu Leu Phe Arg Ala Ile Pro Met Leu Leu Leu Gly Leu Met 1
 5 10 15
 Val Leu Gln Thr Asp Cys Glu Ile Ala Gln Tyr Tyr Ile Asp Glu Glu 20 25 30
 Glu Pro Pro Gly Thr Val Ile Ala Val Leu Ser Gln His Ser Ile Phe 35 40 45
 Asn Thr Thr Asp Ile Pro Ala Thr Asn Phe Arg Leu Met Lys Gln Phe 50 55 60
 Asn Asn Ser Leu Ile Gly Val Arg Glu Ser Asp Gly Gln Leu Ser Ile 65 70 75 80
 Met Glu Arg Ile Asp Arg Glu Gln Ile Cys Arg Gln Ser Leu His Cys 85 90 95
 Asn Leu Ala Leu Asp Val Val Ser Phe Ser Lys Gly His Phe Lys Leu 100 105 110

Leu Asn Val Lys Val Glu Val Arg Asp Ile Asn Asp His Ser Pro His
115 120 125
Phe Pro Ser Glu Ile Met His Val Glu Val Ser Glu Ser Ser Ser Val
130 135 140
Gly Thr Arg Ile Pro Leu Glu Ile Ala Ile Asp Glu Asp Val Gly Ser
145 150 155 160
Asn Ser Ile Gln Asn Phe Gln Ile Ser Asn Asn Ser His Phe Ser Ile
165 170 175
Asp Val Leu Thr Arg Ala Asp Gly Val Lys Tyr Ala Asp Leu Val Leu
180 185 190
Met Arg Glu Leu Asp Arg Glu Ile Gln Pro Thr Tyr Ile Met Glu Leu
195 200 205
Leu Ala Met Asp Gly Gly Val Pro Ser Leu Ser Gly Thr Ala Val Val
210 215 220
Asn Ile Arg Val Leu Asp Phe Asn Asp Asn Ser Pro Val Phe Glu Arg
225 230 235 240
Ser Thr Ile Ala Val Asp Leu Val Glu Asp Ala Pro Leu Gly Tyr Leu
245 250 255
Leu Leu Glu Leu His Ala Thr Asp Asp Asp Glu Gly Val Asn Gly Glu
260 265 270
Ile Val Tyr Gly Phe Ser Thr Leu Ala Ser Gln Glu Val Arg Gln Leu
275 280 285
Phe Lys Ile Asn Ser Arg Thr Gly Ser Val Thr Leu Glu Gly Gln Val
290 295 300
Asp Phe Glu Thr Lys Gln Thr Tyr Glu Phe Glu Val Gln Ala Gln Asp
305 310 315 320
Leu Gly Pro Asn Pro Leu Thr Ala Thr Cys Lys Val Thr Val His Ile
325 330 335
Leu Asp Val Asn Asp Asn Thr Pro Ala Ile Thr Ile Thr Pro Leu Thr
340 345 350
Thr Val Asn Ala Gly Val Ala Tyr Ile Pro Glu Thr Ala Thr Lys Glu
355 360 365
Asn Phe Ile Ala Leu Ile Ser Thr Thr Asp Arg Ala Ser Gly Ser Asn
370 375 380

Gly Gln Val Arg Cys Thr Leu Tyr Gly His Glu His Phe Lys Leu Gln
385 390 395 400
Gln Ala Tyr Glu Asp Ser Tyr Met Ile Val Thr Thr Ser Thr Leu Asp
405 410 415
Arg Glu Asn Ile Ala Ala Tyr Ser Leu Thr Val Val Ala Glu Asp Leu
420 425 430
Gly Phe Pro Ser Leu Lys Thr Lys Lys Tyr Tyr Thr Val Lys Val Ser
435 440 445
Asp Glu Asn Asp Asn Ala Pro Val Phe Ser Lys Pro Gln Tyr Glu Ala
450 455 460
Ser Ile Leu Glu Asn Asn Ala Pro Gly Ser Tyr Ile Thr Thr Val Ile
465 470 475 480
Ala Arg Asp Ser Asp Ser Asp Gln Asn Gly Lys Val Asn Tyr Arg Leu
485 490 495
Val Asp Ala Lys Val Met Gly Gln Ser Leu Thr Thr Phe Val Ser Leu
500 505 510
Asp Ala Asp Ser Gly Val Leu Arg Ala Val Arg Ser Leu Asp Tyr Glu
515 520 525
Lys Leu Lys Gln Leu Asp Phe Glu Ile Glu Ala Ala Asp Asn Gly Ile
530 535 540
Pro Gln Leu Ser Thr Arg Val Gln Leu Asn Leu Arg Ile Val Asp Gln
545 550 555 560
Asn Asp Asn Cys Pro Val Ile Thr Asn Pro Leu Leu Asn Asn Gly Ser
565 570 575
Gly Glu Val Leu Leu Pro Ile Ser Ala Pro Gln Asn Tyr Leu Val Phe
580 585 590
Gln Leu Lys Ala Glu Asp Ser Asp Glu Gly His Asn Ser Gln Leu Phe
595 600 605
Tyr Thr Ile Leu Arg Asp Pro Ser Arg Leu Phe Ala Ile Asn Lys Glu
610 615 620
Ser Gly Glu Val Phe Leu Lys Gln Leu Asn Ser Asp His Ser Glu
625 630 635 640
Asp Leu Ser Ile Val Val Ala Val Tyr Asp Leu Gly Arg Pro Ser Leu
645 650 655

Ser Thr Asn Ala Thr Val Lys Phe Ile Leu Thr Asp Ser Phe Pro Ser
660 665 670
Asn Val Glu Val Val Ile Leu Gln Pro Ser Ala Glu Glu Gln His Gln
675 680 685
Ile Asp Met Ser Ile Ile Phe Ile Ala Val Leu Ala Gly Gly Cys Ala
690 695 700
Leu Leu Leu Ala Ile Phe Phe Val Ala Cys Thr Cys Lys Lys Lys
705 710 715 720
Ala Gly Glu Phe Lys Gln Val Pro Glu Gln His Gly Thr Cys Asn Glu
725 730 735
Glu Arg Leu Leu Ser Thr Pro Ser Pro Gln Ser Val Ser Ser Ser Leu
740 745 750
Ser Gln Ser Glu Ser Cys Gln Leu Ser Ile Asn Thr Glu Ser Glu Asn
755 760 765
Cys Ser Val Ser Ser Asn Gln Glu Gln His Gln Gln Thr Gly Ile Lys
770 775 780
His Ser Ile Ser Val Pro Ser Tyr His Thr Ser Gly Trp His Leu Asp
785 790 795 800
Asn Cys Ala Met Ser Ile Ser Gly His Ser His Met Gly His Ile Ser
805 810 815
Thr Lys Asp Ser Gly Lys Gly Asp Ser Asp Phe Asn Asp Ser Asp Ser
820 825 830
Asp Thr Ser Gly Glu Ser Gln Lys Lys Ser Ile Glu Gln Pro Met Gln
835 840 845
Ala Gln Ala Ser Ala Gln Tyr Thr Asp Glu Ser Ala Gly Phe Arg His
850 855 860
Ala Asp Asn Tyr Phe Ser His Arg Ile Asn Lys Gly Pro Glu Asn Gly
865 870 875 880
Asn Cys Thr Leu Gln Tyr Glu Lys Gly Tyr Arg Leu Ser Tyr Ser Val
885 890 895
Ala Pro Ala His Tyr Asn Thr Tyr His Ala Arg Met Pro Asn Leu His
900 905 910
Ile Pro Asn His Thr Leu Arg Asp Pro Tyr Tyr His Ile Asn Asn Pro
915 920 925

Val Ala Asn Arg Met His Ala Glu Tyr Glu Arg Asp Leu Val Asn Arg
 930 935 940
 Ser Ala Thr Leu Ser Pro Gln Arg Ser Ser Ser Arg Tyr Gln Glu Phe
 945 950 955 960
 Asn Tyr Ser Pro Gln Ile Ser Arg Gln Leu His Pro Ser Glu Ile Ala
 965 970 975
 Thr Thr Phe

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3655 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCCCAG AGATGAACTC CTTGAGATTG TTTTAAATGA CTGCAGGTCT GGAAGGATTC 60
 ACATTGCCAC ACTGTTTCTA GGCATGAAAA AACTGCAAGT TTCAACTTTG TTTTTGGTGC 120
 AACCTTGATT CTTCAAGATG CTGCTTCTCT TCAGAGCCAT TCCAATGCTG CTGTTGGGAC 180
 TGATGGTTTT ACAAAACAGAC TGTGAAATTG CCCAGTACTA CATAGATGAA GAAGAACCCC 240
 CTGGCACTGT AATTGCAGTG TTGTACAAC ACTCCATATT TAACACTACA GATATACCTG 300
 CAACCAATT CCGTCTAATG AAGCAATTAA ATAATTCCCT TATCGGAGTC CGTGAGAGTG 360
 ATGGGCAGCT GAGCATCATG GAGAGGATTG ACCGGGAGCA AATCTGCAGG CAGTCCCTTC 420
 ACTGCAACCT GGCTTTGGAT GTGGTCAGCT TTTCCAAAGG ACACCTCAAG CTTCTGAACG 480
 TGAAAGTGGAA GGTGAGAGAC ATTAATGACC ATAGCCCTCA CTTTCCCAGT GAAATAATGC 540
 ATGTGGAGGT GTCTGAAAGT TCCTCTGTGG GCACCAGGAT TCCTTAGAA ATTGCAATAG 600
 ATGAAGATGT TGGGTCCAAC TCCATCCAGA ACTTCAGAT CTCAAATAAT AGCCACTTCA 660
 GCATTGATGT GCTAACCCAGA GCAGATGGGG TGAAATATGC AGATTTAGTC TTAATGAGAG 720
 AACTGGACAG GGAAATCCAG CCAACATACA TAATGGAGCT ACTAGCAATG GATGGGGGTG 780

TACCATCACT ATCTGGTACT GCAGTGGTTA ACATCCGAGT CCTGGACTTT AATGATAACA	840
GCCCAGTGT TGAGAGAAC ACCATTGCTG TGGACCTAGT AGAGGATGCT CCTCTGGGAT	900
ACCTTTGTT GGAGTTACAT GCTACTGACG ATGATGAAGG AGTGAATGGA GAAATTGTTT	960
ATGGATTCAAG CACTTTGGCA TCTCAAGAGG TACGTCAGCT ATTTAAAATT AACTCCAGAA	1020
CTGGCAGTGT TACTCTTGAA GCCCAAGTTG ATTTGAGAC CAAGCAGACT TACGAATTG	1080
AGGTACAAGC CCAAGATTG GGCCCCAAC CACTGACTGC TACTTGAAA GTAACTGTTC	1140
ATATACTTGA TGTAAATGAT AATACCCAG CCATCACTAT TACCCCTCTG ACTACTGAA	1200
ATGCAGGAGT TGCCTATATT CCAGAAACAG CCACAAAGGA GAACTTTATA GCTCTGATCA	1260
GCACACTGA CAGAGCCTCT GGATCTAATG GACAAGTTCG CTGTACTCTT TATGGACATG	1320
AGCACTTTAA ACTACAGCAA GCTTATGAGG ACAGTTACAT GATAGTTACC ACCTCTACTT	1380
■ TAGACAGGGA AAACATAGCA GCGTACTCTT TGACAGTAGT TGCAGAAGAC CTTGGCTTCC	1440
CCTCATTGAA GACCAAAAG TACTACACAG TCAAGGTTAG TGATGAGAAT GACAATGCAC	1500
CTGTATTTTC TAAACCCAG TATGAAGCTT CTATTCTGGA AAATAATGCT CCAGGCTCTT	1560
ATATAACTAC AGTGATAGCC AGAGACTCTG ATAGTGATCA AAATGGAAA GTAAATTACA	1620
GACTTGTGGA TGCAAAAGTG ATGGGCCAGT CACTAACAAAC ATTTGTTCT CTTGATGCGG	1680
ACTCTGGAGT ATTGAGAGCT GTTAGGTCTT TAGACTATGA AAAACTAAA CAACTGGATT	1740
TTGAAATTGA AGCTGCAGAC AATGGGATCC CTCAACTCTC CACTCGCGTT CAACTAAATC	1800
TCAGAATAGT TGATCAAAAT GATAATTGCC CTGTGATAAC TAATCCTCTT CTTAATAATG	1860
GCTCGGGTGA AGTTCTGCTT CCCATCAGCG CTCCTCAAAA CTATTTAGTT TTCCAGCTCA	1920
AAGCCGAGGA TTCAGATGAA GGGCACAAC CCCAGCTGTT CTATACCATA CTGAGAGATC	1980
CAAGCAGATT GTTGCCATT AACAAAGAAA GTGGTGAAGT GTTCCTGAAA AAACAATTAA	2040
ACTCTGACCA TTCAGAGGAC TTGAGCATAG TAGTTGCAGT GTATGACTTG GGAAGACCTT	2100
CATTATCCAC CAATGCTACA GTTAAATTCA TCCTCACCGA CTCTTTCCCT TCTAACGTTG	2160
AAGTCGTTAT TTTGCAACCA TCTGCAGAAG AGCAGCACCA GATCGATATG TCCATTATAT	2220
TCATTGCAGT GCTGGCTGGT GGTTGTGCTT TGCTACTTTT GGCCATCTT TTTGTGGCCT	2280
GTACTTGAA AAAGAAAGCT GGTGAATTAA AGCAGGTACC TGAACAAACAT GGAACATGCA	2340

ATGAAGAACG CCTGTTAACG ACCCCATCTC CCCAGTCGGT CTCTTCTTCT TTGTCTCAGT 2400
 CTGAGTCATG CCAACTCTCC ATCAATACTG AATCTGAGAA TTGCAGCGTG TCCTCTAAC 2460
 AAGAGCAGCA TCAGCAAACA GGCATAAAAGC ACTCCATCTC TGTACCATCT TATCACACAT 2520
 CTGGTTGGCA CCTGGACAAT TGTGCAATGA GCATAAGTGG ACATTCTCAC ATGGGGCACA 2580
 TTAGTACAAA GGACAGTGGC AAAGGAGATA GTGACTTCAA TGACAGTGAC TCTGATACTA 2640
 GTGGAGAAC ACAAAAGAAC AGCATTGAGC AGCCAATGCA GGCACAAGCC AGTGCTCAAT 2700
 ACACAGATGA ATCAGCAGGG TTCCGACATG CCGATAACTA TTTCAGCCAC CGAATCAACA 2760
 AGGGTCCAGA AAATGGGAAC TGCACATTGC AATATGAAAA GGGCTATAGA CTGTCTTACT 2820
 CTGTAGCTCC TGCTCATTAC AATACCTACC ATGCAAGAAC GCCTAACCTG CACATACCGA 2880
 ACCATACCCCT TAGAGACCCT TATTACCAT A TCAATAATCC TGTTGCTAAT CGGATGCACG 2940
 CGGAATATGA AAGAGATTAA GTCAACAGAA GTGCAACGTT ATCTCCGCAG AGATCGTCTA 3000
 GCAGATACCA AGAATTCAAT TACAGTCCGC AGATATCAAG ACAGCTTCAT CCTTCAGAAA 3060
 TTGCTACAAC CTTTTAATCA TTAGGCATGC AAGTGAGAAT GCACAAAGGC AAGTGCTTTA 3120
 GCATGAAAGC TAAATATATG GAGTCTCCCC TTTCCCTCTG ATGGATGGGG GGAGACACAG 3180
 GACAGTGCAT AAATATACAG CTGCTTCTA TTTGCATTTC ACTTGGGAAT TTTTTGTTTT 3240
 TTTTACATAT TTATTTTCC TGAATTGAAT GTGACATTGT CCTGTCACCT AACTAGCAAT 3300
 TAAATCCACA GACCTACAGT CAAATATTG AGGGCCCCCTG AAACAGCACA TCAGTCAGGA 3360
 CCTAAAGTGG CCTTTTACT TTAGCAGCT CCTGGGTCTG CCCTCTGTGT TAATCAGCCC 3420
 CTGGTCAAGT CCTGAGTAGG ATCATGGCGT TTTTATATGC ATCTCACCTA CTTTGGACGT 3480
 GATTTACACA TAATAGGAAA CGCTTGGTTT CAGTGAAGTC TGTGTTGTAT ATATTCTGTT 3540
 ATATACACGC ATTTGTGTT TGTGTATATA TTTCAAGTCC ATTCAGATAT GTGTATATAG 3600
 TGCAGACCTT GTAAATTAAA TATTCTGATA CTTTTCCCTC AATAAATATT TAAAT 3655

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Val Cys Cys Gly Pro Gly Arg Met Leu Leu Gly Trp Ala Gly Leu
1 5 10 15
Leu Val Leu Ala Ala Leu Cys Leu Leu Gln Val Pro Gly Ala Gln Ala
20 25 30
Ala Ala Cys Glu Pro Val Arg Ile Pro Leu Cys Lys Ser Leu Pro Trp
35 40 45
Asn Met Thr Lys Met Pro Asn His Leu His His Ser Thr Gln Ala Asn
50 55 60
Ala Ile Leu Ala Met Glu Gln Phe Glu Gly Leu Leu Gly Thr His Cys
65 70 75 80
Ser Pro Asp Leu Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys
85 90 95
Thr Ile Asp Phe Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys
100 105 110
Glu Arg Ala Arg Gln Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His
115 120 125
Ser Trp Pro Glu Ser Leu Ala Cys Asp Glu Leu Pro Val Tyr Asp Arg
130 135 140
Gly Val Cys Ile Ser Pro Glu Ala Ile Val Thr Ala Asp Gly Ala Asp
145 150 155 160
Phe Pro Met Asp Ser Ser Thr Gly His Cys Arg Gly Ala Ser Ser Glu
165 170 175
Arg Cys Lys Cys Lys Pro Val Arg Ala Thr Gln Lys Thr Tyr Phe Arg
180 185 190
Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Val Lys Met
195 200 205
Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys
210 215 220
Ala Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Thr
225 230 235 240

Ser Gly Cys Leu Cys Pro Pro Leu Thr Val Asn Glu Glu Tyr Val Ile
 245 250 255
 Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly
 260 265 270
 Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg
 275 280 285
 Trp Asp Met Lys Leu Arg His Leu Gly Leu Gly Lys Thr Asp Ala Ser
 290 295 300
 Asp Ser Thr Gln Asn Gln Lys Ser Gly Arg Asn Ser Asn Pro Arg Pro
 305 310 315 320
 Ala Arg Ser

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2176 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AAGCCTGGGA CCATGGTCTG CTGCGGCCG GGACGGATGC TGCTAGGATG GGCCGGGTTG	60
CTAGTCCTGG CTGCTCTCTG CCTGCTCCAG GTGCCCGGAG CTCAGGCTGC AGCCTGTGAG	120
CCTGTCCGCA TCCCGCTGTG CAAGTCCCTT CCCTGGAACA TGACCAAGAT GCCCAACCAC	180
CTGCAACCACA GCACCCAGGC TAACGCCATC CTGGCCATGG AACAGTTCGA AGGGCTGCTG	240
GGCACCCACT GCAGCCCGGA TCTTCTCTTC TTCCTCTGTG CAATGTACGC ACCCATTGTC	300
ACCATCGACT TCCAGCACGA GCCCATCAAG CCCTGCAAGT CTGTGTGTGA GCGCGCCCGA	360
CAGGGCTGCG AGCCCATTCT CATCAAGTAC CGCCACTCGT GGCCGGAAAG CTTGGCCTGC	420
GACGAGCTGC CGGTGTACGA CCGCGGCGTG TGCATCTCTC CTGAGGCCAT CGTCACCGCG	480
GACGGAGCGG ATTTCCCTAT GGATTCAAGT ACTGGACACT GCAGAGGGC AAGCAGCGAA	540
CGTTGCAAAT GTAAGCCTGT CAGAGCTACA CAGAAGACCT ATTTCCGGAA CAATTACAAC	600

TATGTCATCC GGGCTAAAGT TAAAGAGGTA AAGATGAAAT GTCATGATGT GACCGCCGTT 660
GTGGAAGTGA AGGAAATTCT AAAGGCATCA CTGGTAAACA TTCCAAGGGA CACCGTCAAT 720
CTTTATACCA CCTCTGGCTG CCTCTGTCTT CCACTTACTG TCAATGAGGA ATATGTCATC 780
ATGGGCTATG AAGACGAGGA ACGTTCCAGG TTACTCTTGG TAGAAGGCTC TATAGCTGAG 840
AAGTGGAAAGG ATCGGCTTGG TAAGAAAGTC AAGCGCTGGG ATATGAAACT CCGACACCTT 900
GGACTGGGTA AAACTGATGC TAGCGATTCC ACTCAGAATC AGAAGTCTGG CAGGAACCTCT 960
AATCCCCGGC CAGCACGCAG CTAAATCCTG AAATGTAAAA GGCCACACCC ACGGACTCCC 1020
TTCTAAGACT GGCGCTGGTG GACTAACAAA GGAAAACCGC ACAGTTGTGC TCGTGACCGA 1080
TTGTTTACCG CAGACACCCGC GTGGCTACCG AAGTTACTTC CGGTCCCCCTT TCTCCTGCTT 1140
CTTAATGGCG TGGGGTTAGA TCCTTTAATA TGTTATATAT TCTGTTTCAT CAATCACGTG 1200
GGGACTGTTC TTTTGCAACC AGAATAGTAA ATTAAATATG TTGATGCTAA GGTTTCTGTA 1260
CTGGACTCCC TGGGTTTAAT TTGGTGTCT GTACCCCTGAT TGAGAATGCA ATGTTTCATG 1320
TAAAGAGAGA ATCCTGGTCA TATCTCAAGA ACTAGATATT GCTGTAAGAC AGCCTCTGCT 1380
GCTGCGCTTA TAGTCTTGTG TTTGTATGCC TTTGTCCATT TCCCTCATGC TGTGAAAGTT 1440
ATACATGTTT ATAAAGGTAG AACGGCATT TGAAATCAGA CACTGCACAA GCAGAGTAGC 1500
CCAACACCAAG GAAGCATTG TGAGGAAACG CCACACAGCA TGACTTATT TCAAGATTGG 1560
CAGGCAGCAA AATAAATAGT GTTGGGAGCC AAGAAAAGAA TATTTTGCTT GGTTAAGGGG 1620
CACACTGGAA TCAGTAGCCC TTGAGCCATT AACAGCAGTG TTCTTCTGGC AAGTTTTGA 1680
TTTGTTCATA AATGTATTCA CGAGCATTAG AGATGAACCT ATAACTAGAC ATCTGTTGTT 1740
ATCTCTATAG CTCTGCTTCC TTCTAAATCA AACCCATTGT TGGATGCTCC CTCTCCATTG 1800
ATAAATAAAT TTGGCTTGCT GTATTGGCCA GGAAAAGAAA GTATTAAAGT ATGCATGCAT 1860
GTGCACCAGG GTGTTATTAA ACAGAGGTAT GTAACTCTAT AAAAGACTAT AATTTACAGG 1920
ACACGGAAAT GTGCACATTG GTTTACTTTT TTTCTTCCTT TTGCTTTGGG CTTGTGATT 1980
TGGTTTTGG TGTGTTTATG TCTGTATTAA GGGGGGTGGG TAGGTTAAG CCATTGCACA 2040
TTCAAGTTGA ACTAGATTAG AGTAGACTAG GCTCATTGGC CTAGACATTA TGATTTGAAT 2100

TTGTGTTGTT TAATGCTCCA TCAAGATGTC TAATAAAAGG AATATGGTTG TCAACAGAGA 2160
2176
CGACAAAC AACAAA

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Val Cys Gly Ser Pro Gly Gly Met Leu Leu Leu Arg Ala Gly Leu
1 5 10 15
Leu Ala Leu Ala Ala Leu Cys Leu Leu Arg Val Pro Gly Ala Arg Ala
20 25 30
Ala Ala Cys Glu Pro Val Arg Ile Pro Leu Cys Lys Ser Leu Pro Trp
35 40 45
Asn Met Thr Lys Met Pro Asn His Leu His His Ser Thr Gln Ala Asn
50 55 60
Ala Ile Leu Ala Ile Glu Gln Phe Glu Gly Leu Leu Gly Thr His Cys
65 70 75 80
Ser Pro Asp Leu Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys
85 90 95
Thr Ile Asp Phe Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys
100 105 110
Glu Arg Ala Arg Gln Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His
115 120 125
Ser Trp Pro Glu Asn Leu Ala Cys Glu Glu Leu Pro Val Tyr Asp Arg
130 135 140
Gly Val Cys Ile Ser Pro Glu Ala Ile Val Thr Ala Asp Gly Ala Asp
145 150 155 160
Phe Pro Met Asp Ser Ser Asn Gly Asn Cys Arg Gly Ala Ser Ser Glu
165 170 175

Arg Cys Lys Cys Lys Pro Ile Arg Ala Thr Gln Lys Thr Tyr Phe Arg
 180 185 190
 Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Ile Lys Thr
 195 200 205
 Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys
 210 215 220
 Ser Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Ser
 225 230 235 240
 Ser Gly Cys Leu Cys Pro Pro Leu Asn Val Asn Glu Glu Tyr Ile Ile
 245 250 255
 Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly
 260 265 270
 Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg
 275 280 285
 Trp Asp Met Lys Leu Arg His Leu Gly Leu Ser Lys Ser Asp Ser Ser
 290 295 300
 Asn Ser Asp Ser Thr Gln Ser Gln Lys Ser Gly Arg Asn Ser Asn Pro
 305 310 315 320
 Arg Gln Ala Arg Asn
 325

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1893 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGCGGGAGCGG GCCTTTGGC GTCCACTGCG CGGCTGCACC CTGCCCGCATC TGCCGGGATC	60
ATGGTCTGCG GCAGCCCGGG AGGGATGCTG CTGCTGCAGG CCGGGCTGCT TGCCCTGGCT	120
GCTCTCTGCC TGCTCCGGGT GCCCGGGCT CGGGCTGCAG CCTGTGAGCC CGTCCGCATC	180
CCCCTGTGCA AGTCCCTGCC CTGGAACATG ACTAAGATGC CCAACCACCT GCACCACAGC	240

ACTCAGGCCA ACGCCATCCT GGCCATCGAG CAGTCGAAG GTCTGCTGGG CACCCACTGC	300
AGCCCCGATC TGCTCTTCTT CCTCTGTGCC ATGTACGCGC CCATCTGCAC CATTGACTTC	360
CAGCACGAGC CCATCAAGCC CTGTAAGTCT GTGTGCGAGC GGGCCCGGCA GGGCTGTGAG	420
CCCATACTCA TCAAGTACCG CCACTCGTGG CCGGAGAACCC TGGCCTGCAGA GGAGCTGCCA	480
GTGTACGACA GGGCGTGTG CATCTCTCCC GAGGCCATCG TTACTGCGGA CGGAGCTGAT	540
TTTCCTATGG ATTCTAGTAA CGGAAACTGT AGAGGGCAA GCAGTGAACG CTGTAAATGT	600
AAGCCTATTA GAGCTACACA GAAGACCTAT TTCCGGAACA ATTACAACTA TGTCAATTGG	660
GCTAAAGTTA AAGAGATAAA GACTAAGTGC CATGATGTGA CTGCAGTAGT GGAGGTGAAG	720
GAGATTCTAA AGTCCTCTCT GGTAAACATT CCACGGGACA CTGTCAACCT CTATACCAGC	780
TCTGGCTGCC TCTGCCCTCC ACTTAATGTT AATGAGGAAT ATATCATCAT GGGCTATGAA	840
GATGAGGAAC GTTCCAGATT ACTCTTGGTG GAAGGCTCTA TAGCTGAGAA GTGGAAGGAT	900
CGACTCGGTA AAAAAGTTAA GCGCTGGGAT ATGAAGCTTC GTCATCTTGG ACTCAGTAAA	960
AGTGATTCTA GCAATAGTGA TTCCACTCAG AGTCAGAAGT CTGGCAGGAA CTCGAACCCC	1020
CGGCAAGCAC GCAACTAAAT CCCGAAATAC AAAAGTAAC ACAGTGGACT TCCTATTAAG	1080
ACTTACTTGC ATTGCTGGAC TAGCAAAGGA AAATTGCACT ATTGCACATC ATATTCTATT	1140
GTTTACTATA AAAATCATGT GATAACTGAT TATTACTTCT GTTCTCTTT TGGTTCTGC	1200
TTCTCTCTTC TCTCAACCCC TTTGTAATGG TTTGGGGCA GACTCTTAAG TATATTGTGA	1260
GTTTCTATT TCACTAATCA TGAGAAAAAC TGTTCTTTG CAATAATAAT AAATTAACAA	1320
TGCTGTTACC AGAGCCTCTT TGCTGAGTCT CCAGATGTTA ATTACTTTC TGCACCCCCAA	1380
TTGGGAATGC AATATTGGAT GAAAAGAGAG GTTCTGGTA TTCACAGAAA GCTAGATATG	1440
CCTTAAACAA TACTCTGCCG ATCTAATTAC AGCCTTATTT TTGTATGCCT TTTGGCATT	1500
CTCCTCATGC TTAGAAAGTT CCAAATGTTT ATAAAGGTAA AATGGCAGTT TGAAGTCAAA	1560
TGTCACATAG GCAAAGCAAT CAAGCACCAG GAAGTGTAA TGAGGAAACA ACACCCAAGA	1620
TGAATTATTT TTGAGACTGT CAGGAAGTAA AATAAATAGG AGCTTAAGAA AGAACATTTC	1680
GCCTGATTGA GAAGCACAAC TGAAACCAGT AGCCGCTGGG GTGTTAATGG TAGCATTCTT	1740
CTTTGGCAA TACATTTGAT TTGTTCATGA ATATATTAAT CAGCATTAGA GAAATGAATT	1800

ATAACTAGAC ATCTGCTGTT ATCACCATAG TTTGTTAA TTTGCTTCCT TTTAAATAAA 1860
CCCATTGGTG AAAGTCAAAA AAAAAAAA AAA 1893